

5

1638

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/721,047

DATE: 05/03/2001

TIME: 17:07:54

Input Set : A:\Ocirs330

Output Set: N:\CRF3\05032001\I721047.raw

ENTERED

```

3 110> APPLICANT: TUMER, NILGUN E.
4      WANG, PINGER
6 120> TITLE OF INVENTION: TRANSGENIC PLANTS PRODUCING A PAP II PROTEIN
8 130> FILE REFERENCE: OCIRS 3.3-060 CONT
10 140> CURRENT APPLICATION NUMBER: 09/721,047
11 141> CURRENT FILING DATE: 2000-11-22
13 160> NUMBER OF SEQ ID NOS: 19
15 170> SOFTWARE: PatentIn Ver. 2.1
17 210> SEQ ID NO: 1
18 211> LENGTH: 1379
19 212> TYPE: DNA
20 213> ORGANISM: Phytolacca americana
22 220> FEATURE:
23 221> NAME/KEY: CDS
24 222> LOCATION: (225)..(1163)
26 400> SEQUENCE: 1
27 ctatgaagtc gggtc aaagc atatacaggc tatgcattgt tagaaacatt gatgcctctg 60
29 atcccgataa acaatacaaaa ttagacaata agatgacata caagtaccta aactgtgtat 120
31 gggggagtg aacctcagct gctaaaaaaa cggttgtaaga aaaaaagaaa gttgtgagtt 180
33 aactacaggg cgaaagtatt ggaactagct agtaggaagg gaag atg aag tgg atg 236
34                                     Met Lys Ser Met
35                                     1
37 ctt gtg gtg aca ata tca ata tgg ctc att ctt gca cca act tca act 284
38 Leu Val Val Thr Ile Ser Ile Trp Leu Ile Leu Ala Pro Thr Ser Thr
39 5      10      15      20
41 tgg gct gtg aat aca atc atc tac aat gtt gga agt acc acc att agc 332
42 Trp Ala Val Asn Thr Ile Ile Tyr Asn Val Gly Ser Thr Thr Ile Ser
43      25      30      35
45 aaa tac gcc act ttt ctg aat gat ctt cgt aat gaa gcg aaa gat cca 380
46 Lys Tyr Ala Thr Phe Leu Asn Asp Leu Arg Asn Glu Ala Lys Asp Pro
47      40      45      50
49 agt tta aaa tgc tat gga ata cca atg ctg ccc aat aca aat aca aat 428
50 Ser Leu Lys Cys Tyr Gly Ile Pro Met Leu Pro Asn Thr Asn Thr Asn
51      55      60      65
53 cca aag tac gtg ttg gtt gag ctc caa ggt tca aat aaa aaa acc atc 476
54 Pro Lys Tyr Val Leu Val Glu Leu Gln Gly Ser Asn Lys Lys Thr Ile
55      70      75      80
57 aca cta atg ctg aga cga aac aat ttg tat gtg atg ggt tat tct gat 524
58 Thr Leu Met Leu Arg Arg Asn Asn Leu Tyr Val Met Gly Tyr Ser Asp
59 85      90      95      100
61 ccc ttt gaa acc aat aaa tgt cgt tac cat atc ttt aat gat atc tca 572
62 Pro Phe Glu Thr Asn Lys Cys Arg Tyr His Ile Phe Asn Asp Ile Ser
63      105      110      115
65 ggt act gaa cgc caa gat gta gag act act ctt tgc cca aat gcc aat 620
66 Gly Thr Glu Arg Gln Asp Val Glu Thr Thr Leu Cys Pro Asn Ala Asn
67      120      125      130
69 tct cgt gtt agt aaa aac ata aac ttt gat agt cga tat cca aca ttg 668

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/721,047

DATE: 05/03/2001

TIME: 17:07:54

Input Set : A:\Ocirs330

Output Set : N:\CRF3\05032001\I721047.raw

```

70 Ser Arg Val Ser Lys Asn Ile Asn Phe Asp Ser Arg Tyr Pro Thr Leu
71      135      140      145
73 gaa tca aaa gcg gga gta aaa tca aga agt cag gtc caa ctg gga att 716
74 Glu Ser Lys Ala Gly Val Lys Ser Arg Ser Gln Val Gln Leu Gly Ile
75      150      155      160
77 caa ata ctc gac agt aat att gga aag att tct gga gtg atg tca ttc 764
78 Gln Ile Leu Asp Ser Asn Ile Gly Lys Ile Ser Gly Val Met Ser Phe
79 165      170      175      180
81 act gag aaa acc gaa gcc gaa ttc cta ttg gta gcc ata caa atg gta 812
82 Thr Glu Lys Thr Glu Ala Glu Phe Leu Leu Val Ala Ile Gln Met Val
83      185      190      195
85 tca gag gca gca aga ttc aag tac ata gag aat cag gtg aaa act aat 860
86 Ser Glu Ala Ala Arg Phe Lys Tyr Ile Glu Asn Gln Val Lys Thr Asn
87      200      205      210
89 ttt aac aga gca ttc aac cct aat ccc aaa gta ctt aat ttg caa gag 908
90 Phe Asn Arg Ala Phe Asn Pro Asn Pro Lys Val Leu Asn Leu Gln Glu
91      215      220      225
93 aca tgg ggt aag att tca aca gca att cat gat gcc aag aat gga gtt 956
94 Thr Trp Gly Lys Ile Ser Thr Ala Ile His Asp Ala Lys Asn Gly Val
95      230      235      240
97 tta ccc aaa cct ctc gag cta gtg gat gcc agt ggt gcc aag tgg ata 1004
98 Leu Pro Lys Pro Leu Glu Leu Val Asp Ala Ser Gly Ala Lys Trp Ile
99 245      250      255      260
101 gtg ttg aga gtg gat gaa atc aag cct gat gta gca ctc tta aac tac 1052
102 Val Leu Arg Val Asp Glu Ile Lys Pro Asp Val Ala Leu Leu Asn Tyr
103      265      270      275
105 gtt ggt ggg agc tgt cag aca act tat aac caa aat gcc atg ttt cct 1100
106 Val Gly Gly Ser Cys Gln Thr Thr Tyr Asn Gln Asn Ala Met Phe Pro
107      280      285      290
109 caa ctt ata atg tct act tat tat aat tac atg gtt aat ctt ggt gat 1148
110 Gln Leu Ile Met Ser Thr Tyr Tyr Asn Tyr Met Val Asn Leu Gly Asp
111      295      300      305
113 cta ttt gaa gga ttc tgatcataaa cataataagg agtatatata tattactcca 1203
114 Leu Phe Glu Gly Phe
115      310
117 actatattat aaagcttaaa taagaggccg tgtaattag tacttggtgc cttttgcttt 1263
119 atgggtgtgt ttattatgcc ttgtatgctt gtaatattat ctagagaaca agatgtactg 1323
121 tgtaatagtc ttgtttgaaa taaaacttcc aattatgatg caaaaaaaaa aaaaaa 1379
124 <210> SEQ ID NO: 2
125 <211> LENGTH: 313
126 <212> TYPE: PRT
127 <213> ORGANISM: Phytolacca americana
129 <400> SEQUENCE: 2
130 Met Lys Ser Met Leu Val Val Thr Ile Ser Ile Trp Leu Ile Leu Ala
131 1 5 10 15
133 Pro Thr Ser Thr Trp Ala Val Asn Thr Ile Ile Tyr Asn Val Gly Ser
134 20 25 30
136 Thr Thr Ile Ser Lys Tyr Ala Thr Phe Leu Asn Asp Leu Arg Asn Glu
137 35 40 45

```

RAW SEQUENCE LISTING

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,047

TIME: 17:07:54

Input Set : A:\Ocirs330

Output Set: N:\CRF3\05032001\I721047.raw

```

139 Ala Lys Asp Pro Ser Leu Lys Cys Tyr Gly Ile Pro Met Leu Pro Asn
140      50      55      60
142 Thr Asn Thr Asn Pro Lys Tyr Val Leu Val Glu Leu Gln Gly Ser Asn
143      65      70      75      80
145 Lys Lys Thr Ile Thr Leu Met Leu Arg Arg Asn Asn Leu Tyr Val Met
146      85      90      95
148 Gly Tyr Ser Asp Pro Phe Glu Thr Asn Lys Cys Arg Tyr His Ile Phe
149      100     105     110
151 Asn Asp Ile Ser Gly Thr Glu Arg Gln Asp Val Glu Thr Thr Leu Cys
152      115     120     125
154 Pro Asn Ala Asn Ser Arg Val Ser Lys Asn Ile Asn Phe Asp Ser Arg
155      130     135     140
157 Tyr Pro Thr Leu Glu Ser Lys Ala Gly Val Lys Ser Arg Ser Gln Val
158      145     150     155     160
160 Gln Leu Gly Ile Gln Ile Leu Asp Ser Asn Ile Gly Lys Ile Ser Gly
161      165     170     175
163 Val Met Ser Phe Thr Glu Lys Thr Glu Ala Glu Phe Leu Leu Val Ala
164      180     185     190
166 Ile Gln Met Val Ser Glu Ala Ala Arg Phe Lys Tyr Ile Glu Asn Gln
167      195     200     205
169 Val Lys Thr Asn Phe Asn Arg Ala Phe Asn Pro Asn Pro Lys Val Leu
170      210     215     220
172 Asn Leu Gln Glu Thr Trp Gly Lys Ile Ser Thr Ala Ile His Asp Ala
173      225     230     235     240
175 Lys Asn Gly Val Leu Pro Lys Pro Leu Glu Val Asp Ala Ser Gly
176      245     250     255
178 Ala Lys Trp Ile Val Leu Arg Val Asp Glu Ile Lys Pro Asp Val Ala
179      260     265     270
181 Leu Leu Asn Tyr Val Gly Gly Ser Cys Gln Thr Thr Tyr Asn Gln Asn
182      275     280     285
184 Ala Met Phe Pro Gln Leu Ile Met Ser Thr Tyr Tyr Asn Tyr Met Val
185      290     295     300
187 Asn Leu Gly Asp Leu Phe Glu Gly Phe
188      305     310
191 <210> SEQ ID NO: 3
192 <211> LENGTH: 934
193 <212> TYPE: DNA
194 <213> ORGANISM: Phytolacca americana
196 <220> FEATURE:
197 <221> NAME/KEY: CDS
198 <222> LOCATION: (1)..(930)
200 <400> SEQUENCE: 3
201 atg aag atg aag gtg tta gaa gta gtt ggg ttg gca ata tcg ata tgg 48
202 Met Lys Met Lys Val Leu Glu Val Val Gly Leu Ala Ile Ser Ile Trp
203      1      5      10      15
205 ctg atg ctt aca cca cca gct tct tca aac ata gtg ttt gac gtt gag 96
206 Leu Met Leu Thr Pro Pro Ala Ser Ser Asn Ile Val Phe Asp Val Glu
207      20      25      30
209 aat gcc aca cca gaa acc tac tct aat ttt ctg act agt ttg cga gaa 144

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/721,047

DATE: 05/03/2001

TIME: 17:07:54

Input Set : A:\Ocirs330

Output Set N:\CRF3\05032001\I721047.raw

```

210 Asn Ala Thr Pro Glu Thr Tyr Ser Asn Phe Leu Thr Ser Leu Arg Glu
211          35          40          45
213 gct gtg aaa gac aag aaa ttg aca tgc cat gga atg ata atg gcc aca 192
214 Ala Val Lys Asp Lys Lys Leu Thr Cys His Gly Met Ile Met Ala Thr
215          50          55          60
217 acc ctc act gaa caa ccc aag tat gtg ttg gtt gac ctc aaa ttc gga 240
218 Thr Leu Thr Glu Gln Pro Lys Tyr Val Leu Val Asp Leu Lys Phe Gly
219 65          70          75          80
221 tct gga aca ttc aca tta gca atc aga agg gga aac tta tat ttg gag 288
222 Ser Gly Thr Phe Thr Leu Ala Ile Arg Arg Gly Asn Leu Tyr Leu Glu
223          85          90          95
225 ggc tat tct gac att tac aat gga aaa tgt cgt tat cgg atc ttc aag 336
226 Gly Tyr Ser Asp Ile Tyr Asn Gly Lys Cys Arg Tyr Arg Ile Phe Lys
227          100          105          110
229 gat tca gaa tcc gat gcc caa gag acc gtt tgc ccc ggg gac aaa agc 384
230 Asp Ser Glu Ser Asp Ala Gln Glu Thr Val Cys Pro Gly Asp Lys Ser
231          115          120          125
233 aag cct ggc act cag aat aat atc ccc tat gaa aag agt tac aaa ggg 432
234 Lys Pro Gly Thr Gln Asn Asn Ile Pro Tyr Glu Lys Ser Tyr Lys Gly
235          130          135          140
237 atg gaa tca aag ggt ggg gct aga act aaa tta ggg tta gga aag ata 480
238 Met Glu Ser Lys Gly Gly Ala Arg Thr Lys Leu Gly Leu Gly Lys Ile
239 145          150          155          160
241 aca ctc aag agt cga atg ggt aaa atc tac ggc aag gat gca acg gat 528
242 Thr Leu Lys Ser Arg Met Gly Lys Ile Tyr Gly Lys Asp Ala Thr Asp
243          165          170          175
245 cag aag cag tat caa aaa aat gag gct gaa ttt ctt ctt ata gcc gtt 576
246 Gln Lys Gln Tyr Gln Lys Asn Glu Ala Glu Phe Leu Leu Ile Ala Val
247          180          185          190
249 caa atg gtt act gag gca tca agg ttc aaa tac att gag aac aaa gtg 624
250 Gln Met Val Thr Glu Ala Ser Arg Phe Lys Tyr Ile Glu Asn Lys Val
251          195          200          205
253 aag gct aaa ttt gat gat gcc aat ggg tat cag cca gat cct aaa gct 672
254 Lys Ala Lys Phe Asp Asp Ala Asn Gly Tyr Gln Pro Asp Pro Lys Ala
255          210          215          220
257 att tcc cta gag aaa aat tgg gac agt gtt tct aag gtc att gca aaa 720
258 Ile Ser Leu Glu Lys Asn Trp Asp Ser Val Ser Lys Val Ile Ala Lys
259 225          230          235          240
261 gtt ggc acc tcc ggt gat agt act gtt act tta cct gga gac cta aaa 768
262 Val Gly Thr Ser Gly Asp Ser Thr Val Thr Leu Pro Gly Asp Leu Lys
263          245          250          255
265 gat gag aat aat aaa cct tgg act acg gcc acc atg aac gac ctt aag 816
266 Asp Glu Asn Asn Lys Pro Trp Thr Thr Ala Thr Met Asn Asp Leu Lys
267          260          265          270
269 aac gac att atg gca ctc cta acc cac gtt act tgc aag gtt aaa agt 864
270 Asn Asp Ile Met Ala Leu Leu Thr His Val Thr Cys Lys Val Lys Ser
271          275          280          285
273 tcc atg ttc cct gaa att atg tcc tat tat tat agg act agt att agt 912
274 Ser Met Phe Pro Glu Ile Met Ser Tyr Tyr Tyr Arg Thr Ser Ile Ser

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/721,047

DATE: 05/03/2001

TIME: 17:07:54

Input Set : A:\Ocirs330

Output Set: N:\CRF3\05032001\I721047.raw

```

275      290      295      300
277 aac att ggt gaa ttc gag tgat          934
278 Asn Leu Gly Glu Phe Glu
279 305      310
282 <210> SEQ ID NO: 4
283 <211> LENGTH: 310
284 <212> TYPE: PRT
285 <213> ORGANISM: Phytolacca americana
287 <400> SEQUENCE: 4
288 Met Lys Met Lys Val Leu Glu Val Val Gly Leu Ala Ile Ser Ile Trp
289   1      5      10      15
291 Leu Met Leu Thr Pro Pro Ala Ser Ser Asn Ile Val Phe Asp Val Glu
292      20      25      30
294 Asn Ala Thr Pro Glu Thr Tyr Ser Asn Phe Leu Thr Ser Leu Arg Glu
295      35      40      45
297 Ala Val Lys Asp Lys Lys Leu Thr Cys His Gly Met Ile Met Ala Thr
298      50      55      60
300 Thr Leu Thr Glu Gln Pro Lys Tyr Val Leu Val Asp Leu Lys Phe Gly
301   65      70      75      80
303 Ser Gly Thr Phe Thr Leu Ala Ile Arg Arg Gly Asn Leu Tyr Leu Glu
304      85      90      95
306 Gly Tyr Ser Asp Ile Tyr Asn Gly Lys Cys Arg Tyr Arg Ile Phe Lys
307      100     105     110
309 Asp Ser Glu Ser Asp Ala Gln Glu Thr Val Cys Pro Gly Asp Lys Ser
310     115     120     125
312 Lys Pro Gly Thr Gln Asn Asn Ile Pro Tyr Glu Lys Ser Tyr Lys Gly
313     130     135     140
315 Met Glu Ser Lys Gly Gly Ala Arg Thr Lys Leu Gly Leu Gly Lys Ile
316 145     150     155     160
318 Thr Leu Lys Ser Arg Met Gly Lys Ile Tyr Gly Lys Asp Ala Thr Asp
319     165     170     175
321 Gln Lys Gln Tyr Gln Lys Asn Glu Ala Glu Phe Leu Leu Ile Ala Val
322     180     185     190
324 Gln Met Val Thr Glu Ala Ser Arg Phe Lys Tyr Ile Glu Asn Lys Val
325     195     200     205
327 Lys Ala Lys Phe Asp Asp Ala Asn Gly Tyr Gln Pro Asp Pro Lys Ala
328     210     215     220
330 Ile Ser Leu Glu Lys Asn Trp Asp Ser Val Ser Lys Val Ile Ala Lys
331 225     230     235     240
333 Val Gly Thr Ser Gly Asp Ser Thr Val Thr Leu Pro Gly Asp Leu Lys
334     245     250     255
336 Asp Glu Asn Asn Lys Pro Trp Thr Thr Ala Thr Met Asn Asp Leu Lys
337     260     265     270
339 Asn Asp Ile Met Ala Leu Leu Thr His Val Thr Cys Lys Val Lys Ser
340     275     280     285
342 Ser Met Phe Pro Glu Ile Met Ser Tyr Tyr Tyr Arg Thr Ser Ile Ser
343     290     295     300
345 Asn Leu Gly Glu Phe Glu
346 305     310

```

VERIFICATION SUMMARY

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,047

TIME: 17:07:55

Input Set : A:\Ocirs330

Output Set: N:\CRF3\05032001\I721047.raw